

A limit process for a sequence of partial sums of residuals of a simple regression against order statistics with Markov-modulated noise

Artyom Kovalevskii ^{*}, Evgeny Shatalin [†]

Abstract

We consider a simple regression model where a regressor is composed of order statistics and a noise is Markov-modulated. We introduce an empirical bridge of regression residuals and prove its weak convergence to a centered Gaussian process.

Keywords: simple regression model; order statistics; Markov-modulated noise; regression residuals; empirical bridge.

1 Introduction and main results

Brown et al. (1975) proposed a test for change of regression at unknown time. Their approach is based on computation of recursive residuals. MacNeill (1978) studied a linear regression against values of continuously differentiable functions. He obtained limit processes for sequences of partial sums of regression residuals. Later Bischoff (1997) showed that the MacNeill's theorem holds in more general setting, namely for continuous regressor functions. Aue et al. (2008) introduced a new test for polynomial regression functions which is analogous to the classical likelihood test. Stute (1997) proposed a class of tests that are based on regression residuals. His general approach also allows to analyse models with order statistics regressors.

We consider another model of a simple linear regression against order statistics where the noise is Markov-modulated, and analyse a limit process for sums of regression residuals.

To define the model, we introduce 3 mutually independent families of random variables:

^{*}E-mail: pandorra@ngs.ru

[†]E-mail: sh_e_v_89@list.ru

1) $\{\varepsilon_i^v, i \geq 1, 1 \leq v \leq M\}$, a family of independent random variables where $\{\varepsilon_i^v, i \geq 1\}$ are identically distributed for each v , $\mathbf{E}\varepsilon_1^v = 0$, $\mathbf{D}\varepsilon_1^v = \sigma_v^2 \geq 0$ and $\sum_{v=1}^M \sigma_v^2 > 0$;

2) $\{\xi_i\}_{i=1}^\infty$, a sequence of i.i.d. random variables with distribution function F and finite positive variance $\mathbf{Var}\xi$;

3) $\{V_i\}_{i=1}^\infty$, an irreducible aperiodic Markov chain on the state space $\{1, \dots, M\}$ with stationary distribution $\{\pi_i\}_{i=1}^M$.

For any $n = 1, 2, \dots$, let $X_{ni} = \xi_{i:n}$ be the i -th order statistic of the first n random variables ξ_1, \dots, ξ_n , where, in particular, $X_{n1} = \min_{1 \leq i \leq n} \xi_i$ and $X_{nn} = \max_{1 \leq i \leq n} \xi_i$.

In this article, we consider the following regression model:

$$Y_{ni} = a + bX_{ni} + \varepsilon_i^{V_i}, \quad n \geq 1, i = 1, \dots, n.$$

For this model, we introduce an *empirical bridge* and show its weak convergence to a centered Gaussian process.

The novelty of our model lies in consideration both ordered regressors and Markov-modulated noise.

Let

$$\hat{b}_n = \frac{\overline{XY} - \overline{X} \overline{Y}}{\overline{X^2} - \overline{X}^2}, \quad \hat{a}_n = \overline{Y} - \hat{b}_n \overline{X}.$$

be the classical Gauss-Markov estimators for a and b .

Define *fitted values* $\{\hat{Y}_{ni}\}$, *regression residuals* $\{\hat{\varepsilon}_{ni}\}$ and their *partial sums* $\{\hat{\Delta}_{ni}^0\}$, by $\hat{Y}_{ni} = \hat{a}_n + \hat{b}_n X_{ni}$, $\hat{\varepsilon}_{ni} = Y_{ni} - \hat{Y}_{ni}$ and $\hat{\Delta}_{ni}^0 = \hat{\varepsilon}_{n1} + \dots + \hat{\varepsilon}_{ni}$, for $1 \leq i \leq n$, $n \geq 1$.

In what follows, we write for short: $Y_{ni} = Y_i$, $X_{ni} = X_i$, $\hat{Y}_{ni} = \hat{Y}_i$, $\hat{\varepsilon}_{ni} = \hat{\varepsilon}_i$ and $\hat{\Delta}_{ni}^0 = \hat{\Delta}_i^0$.

A *random polygon* Z_n is a piecewise linear function with nodes $(k/n, \hat{\Delta}_k^0/\sigma\sqrt{n})$, for $k = 0, \dots, n$.

Further, an *empirical bridge* is a random polygon \hat{Z}_n with nodes $(k/n, (\hat{\Delta}_k^0 - k/n\hat{\Delta}_n^0)/\sqrt{\hat{\sigma}^2 n})$ where $\hat{\sigma}^2 = \overline{\hat{\varepsilon}^2} - (\overline{\hat{\varepsilon}})^2$ is an estimator of variance $\sigma^2 = \sum_{v=1}^M \sigma_v^2 \pi_v$.

Let $GL_F(t) = \int_0^t F^{-1}(s) ds$ be the *theoretical general Lorenz curve* (Gastwirth, 1971; Davydov and Zitikis, 2004) where $F^{-1}(s) = \sup\{x : F(x) < s\}$ is the inverse of distribution function $F(x)$. Let $GL_F^0(t) = GL_F(t) - tGL_F(1)$ be its centered version. Similarly, let $GL_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} \xi_{i:n}$ be the *empirical Lorenz curve*. Goldie (1977) showed that, as $n \rightarrow \infty$, the empirical Lorenz curve converges a.s. to the theoretical curve in the uniform metric, i.e. $\sup_{t \in \mathbf{R}} |GL_n(t) - GL_F(t)| \rightarrow 0$ a.s.

Now we formulate the main result of the paper.

Theorem 1 *Both the random polygon Z_n and the empirical bridge \hat{Z}_n converge weakly, as $n \rightarrow \infty$, to the centered Gaussian process Z_F^0 with covariance*

kernel, $K_F^0(t, s)$, given by

$$K_F^0(t, s) = \min\{t, s\} - ts - \frac{GL_F^0(t)GL_F^0(s)}{\mathbf{Var}\xi_1}, \quad t, s \in [0, 1].$$

Here weak convergence holds in the space $C(0, 1)$ of continuous functions on $[0, 1]$ endowed by the uniform metric.

When the Markov chain degenerates, our model is a very particular case of Stute (1997). Kovalevskii (2013) used this particular model to analyse dependence of a car price on a production year.

In what follows, notation $\xrightarrow{\mathbf{P}}$ states for convergence in probability.

2 Proof of Theorem 1

Let $X_i^0 = X_i - \bar{X}$, $\varepsilon_i^0 = \varepsilon_i^{V_i} - \bar{\varepsilon}$ where $\bar{\varepsilon} = \sum_{i=1}^n \varepsilon_i^{V_i}$.

The proof includes five steps. In the first step, we show that, in the formulae under consideration, the sum $\sum_{i=1}^n \frac{\varepsilon_i^0 X_i^0}{\sqrt{n}}$ may be replaced by the sum $\sum_{i=1}^n \frac{\varepsilon_i^0 \mathbf{E}X_i^0}{\sqrt{n}}$. Secondly, we prove weak convergence of a normalized vector with coordinates $(\hat{\Delta}_{k_1}^0, \dots, \hat{\Delta}_{k_m}^0)$ to a normalized vector with coordinates $(\Delta_{k_1}^0, \dots, \Delta_{k_m}^0)$ where $\Delta_{k_i}^0$ are defined below. Then we prove weak convergence of finite-dimensional distributions. The fourth step contains a proof of relative compactness of the family $\{Z_n(t), 0 \leq t \leq 1\}$. We complete with a proof of convergence of sample variance $\widehat{\sigma}^2$ to variance σ^2 .

Step 1

Note that

$$\hat{\Delta}_k^0 = \sum_{i=1}^k \left(\varepsilon_i^0 - \frac{\overline{X^0 \varepsilon^0}}{\overline{(X^0)^2}} X_i^0 \right). \quad (1)$$

We show that

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n \varepsilon_i^0 X_i^0 - \sum_{i=1}^n \varepsilon_i^0 \mathbf{E}X_i^0 \right) \xrightarrow{\mathbf{P}} 0. \quad (2)$$

Indeed,

$$\begin{aligned} \mathbf{P} \left\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i^0 (X_i^0 - \mathbf{E}X_i^0) \right| \geq \delta \right\} &\leq \frac{\mathbf{Var} \sum_{i=1}^n \varepsilon_i^0 (X_i^0 - \mathbf{E}X_i^0)}{n\delta^2} \\ &= \frac{\sum_{i=1}^n \mathbf{Var}_{\varepsilon_i^{V_i}} \mathbf{Var} X_i^0}{n\delta^2} - \frac{2 \sum_{i,j=1}^n \mathbf{Var}_{\varepsilon_i^{V_i}} \mathbf{cov}(X_i^0, X_j^0)}{n^2\delta^2} + \frac{\sum_{k=1}^n \mathbf{Var}_{\varepsilon_k^{V_k}} \sum_{i,j=1}^n \mathbf{cov}(X_i^0, X_j^0)}{n^4\delta^2}. \end{aligned}$$

The sum of covariances admits the follows upper bound.

$$\begin{aligned} \sum_{i=1}^n 2|\mathbf{cov}(X_i, \bar{X})| &\leq \sum_{i=1}^n 2\sqrt{\mathbf{Var}X_i \mathbf{Var}\bar{X}} \leq \sum_{i=1}^n 2\left(\frac{1 + \mathbf{Var}X_i}{2}\right) \sqrt{\frac{\mathbf{Var}X_i}{n}} \\ &= \sqrt{\mathbf{Var}X_i} \left(\sqrt{n} + \sum_{i=1}^n \mathbf{Var}X_i\right). \end{aligned}$$

Theorem 1 (Hoeffding, 1953) implies $\frac{1}{n} \sum_{i=1}^n \mathbf{Var}X_i \rightarrow 0$ as $n \rightarrow \infty$. Therefore $\frac{1}{n} \sum_{i=1}^n \mathbf{Var}X_i^0 \rightarrow 0$. Notice also that $\frac{1}{n} \mathbf{Var} \sum_{i=1}^n \varepsilon_i^{V_i} \rightarrow 0$ as $n \rightarrow 0$, so (2) follows.

Step 2 Let $[t]$ be the integer part of t . For any fixed m and for $0 \leq s_1 < \dots < s_m \leq 1$, $k_i = [ns_i]$, we establish weak convergence, as $n \rightarrow \infty$, of vector $\vec{\eta} = \frac{1}{\sigma\sqrt{n}}(\hat{\Delta}_{k_1}^0, \dots, \hat{\Delta}_{k_m}^0)$ to vector $\vec{Z}_F^0 = (Z_F^0(s_1), \dots, Z_F^0(s_m))$.

From (1), (2) and from convergences $\overline{(X^0)^2} \rightarrow \mathbf{Var}\xi_1$ a.s., $\frac{1}{n} \sum_{i=1}^{k_i} X_i^0 \rightarrow GL_F^0(s_i)$ a.s. (Goldie, 1975), it is enough to prove $\vec{\zeta} \Rightarrow \vec{Z}_F^0$ where $\vec{\zeta} = \frac{1}{\sigma\sqrt{n}}(\Delta_{k_1}^0, \dots, \Delta_{k_m}^0)$,

$$\Delta_{k_j}^0 = \sum_{i=1}^{k_j} \varepsilon_i^0 - \frac{GL_F^0(s_j)}{\mathbf{Var}\xi_1} \sum_{i=1}^n \varepsilon_i^0 \mathbf{E}X_i^0 = \sum_{i=1}^{k_j} \varepsilon_i^0 - \frac{GL_F^0(s_j)}{\mathbf{Var}\xi_1} \sum_{i=1}^n \varepsilon_i^{V_i} \mathbf{E}X_i^0.$$

Step 3 We prove weak convergence $\vec{\zeta} \Rightarrow \vec{Z}_F^0$ using characteristic functions. Notice that

$$\begin{aligned} &\sum_{j=1}^m t_j \left(\sum_{i=1}^{k_j} (\varepsilon_i^{V_i} - \bar{\varepsilon}) - \frac{GL_F^0(s_j)}{\mathbf{Var}\xi_1} \sum_{i=1}^n \varepsilon_i^{V_i} \mathbf{E}X_i^0 \right) \\ &= \sum_{i=1}^n \varepsilon_i^{V_i} \sum_{j=1}^m t_j \left(\mathbf{I}\{i \leq k_j\} - \frac{k_j}{n} - \frac{GL_F^0(s_j)}{\mathbf{Var}\xi_1} \mathbf{E}X_i^0 \right). \end{aligned}$$

It is well known that the finiteness of $\mathbf{E}\psi_1$ implies convergence $\frac{\psi_{n:n}}{n} \rightarrow 0$ a.s. and in mean for a sequence of i.i.d random variables $\psi_1, \dots, \psi_n, \dots$ and, more generally, for a stationary ergodic sequence as a consequence of the subadditive ergodic theorem (Kingman, 1968).

Applying this fact and using Hölder's inequality we have $\mathbf{E}X_i^0 = o(\sqrt{n})$ uniformly in $1 \leq i \leq n$.

Let $\beta_i = \sum_{j=1}^m t_j \left(\mathbf{I}\{i \leq k_j\} - \frac{k_j}{n} - \frac{GL_F^0(s_j)}{\mathbf{D}\xi_1} \mathbf{E}X_i^0 \right)$. Then $\sum_{i=1}^n \frac{\beta_i^2}{n} \rightarrow C_F := \sum_{j_1=1}^m \sum_{j_2=1}^m t_{j_1} t_{j_2} K_F^0(s_{j_1}, s_{j_2})$ a.s. and characteristic function $\varphi_{\vec{\zeta}}(\vec{t})$ converges to $\exp(-C_F/2)$ a.s. Then convergence of finite-dimensional distributions follows.

Step 4. We show that

the family of distributions $\{Z_n(t), 0 \leq t \leq 1\}$ is relatively compact. (3)

Let $S_k = \sum_{i=1}^k \xi_{i:n}$, $k = 1, \dots, n$, $S_0 = 0$.

By Prokhorov's theorem (section 1 §6 in Billingsley, 1968) it suffices to show that the family of distributions of random processes $\left\{ \frac{\hat{\Delta}_{[nt]}^0}{\sigma\sqrt{n}}, 0 \leq t \leq 1 \right\}$, $n = 1, 2, \dots$, is tight. Put $k = [nt]$ and let

$$\hat{\Delta}_k = \sum_{i=1}^k \left(\varepsilon_i^{V_i} - \frac{\overline{X^0 \varepsilon^0}}{(X^0)^2} X_i \right).$$

Then $\hat{\Delta}_k^0 = \hat{\Delta}_k - \frac{k}{n} \hat{\Delta}_n$. The invariance principle (e.g., part 1 of chapter 19 in Borovkov, 1998) implies tightness of the family $\left\{ \frac{\sum_{i=1}^k \varepsilon_i^{V_i}}{\sigma\sqrt{n}}, 0 \leq t \leq 1 \right\}$. So, to show (3), it is enough to establish tightness of

$$\left\{ \frac{\overline{X^0 \varepsilon^0} \sqrt{n}}{\sigma(X^0)^2} \frac{S_k}{n}, 0 \leq t \leq 1 \right\}.$$

In turn, by Theorem 8.3 (Billingsley, 1968), it suffices to prove that, for any $\varepsilon > 0$, $\alpha > 0$, there are $0 < \delta < 1$, $n_0 \in \mathbf{N}$ such that

$$\frac{1}{\delta} \mathbf{P} \left\{ \sup_{t \leq s \leq t+\delta} \left| \frac{\overline{X^0 \varepsilon^0} \sqrt{n}}{\sigma(X^0)^2} \frac{S_{[ns]} - S_{[nt]}}{n} \right| \geq \varepsilon \right\} \leq \alpha, \quad (4)$$

for all $n > n_0$, $0 \leq t \leq 1$.

Notice that $\frac{\overline{X^0 \varepsilon^0} \sqrt{n}}{\sigma(X^0)^2} \implies \frac{\zeta}{\sqrt{\mathbf{Var} \xi_1}}$, and (Goldie, 1977)

$$\sup_{t \leq s \leq t+\delta} \left| \frac{S_{[ns]} - S_{[nt]}}{n} \right| \rightarrow \sup_{t \leq s \leq t+\delta} |GL_F(s) - GL_F(t)| \text{ a.s..}$$

Here ζ is a standard normal random variable and $GL_F(x)$ is the general Lorenz curve.

By Cauchy-Bunyakowsky inequality,

$$\sup_{t \leq s \leq t+\delta} |GL_F(s) - GL_F(t)| \leq \sup_{t \leq s \leq t+\delta} \int_t^s |F^{-1}(x)| dx \leq \sqrt{\delta \mathbf{E} \xi_1^2}.$$

So one may choose a positive δ that satisfies (4).

Step 5. It remains to prove $\widehat{\sigma^2} \xrightarrow{\mathbf{P}} \sigma^2$. Indeed,

$$\widehat{\varepsilon^2} = \frac{1}{n} \sum_{i=1}^n \left(\varepsilon_i^{V_i} - \bar{\varepsilon} - \frac{\overline{X^0 \varepsilon^0}}{(X^0)^2} (X_i - \bar{X}) \right)^2 = (\bar{\varepsilon}^0)^2 - \frac{(\overline{X^0 \varepsilon^0})^2}{(X^0)^2} \xrightarrow{\mathbf{P}} \sigma^2.$$

This completes the proof of Theorem 1.

Acknowledgements

The work was supported in part by Russian Foundation of Basic Researches (grant 13-01-00661), Novosibirsk State Technical University (project 2.1.1-2013), Government of Novosibirsk Region, Russia.

Authors would like to thank Sergey Foss for many useful discussions, Evgeny Baklanov, Igor Borisov, Alexander Sakhanenko and Yuliana Linke for useful references.

References

- ARNOLD, J. R., LIBBY, W. F., 1949. Age determinations by radiocarbon content: checks with samples of known age. *Science* 110, 678–680.
- AUE, A., HORVATH, L., HUSKOVA, M., KOKOSZKA, P., 2008. Testing for change in polynomial regression. *Bernoulli* 14, 637–660.
- BILLINGSLEY, P., 1968. *Convergence of Probability Measures*. New York: John Wiley & Sons.
- BISCHOFF, W., 1998. A functional central limit theorem for regression models. *Ann. Stat.* 26, 1398–1410.
- BOROVKOV, A. A., 1998. *Probability theory*. Amsterdam: Gordon and Breach Science Publishers.
- BROWN, R. L., DURBIN, J., EVANS, J. M., 1975. Techniques for testing the constancy of regression relationships over time. *J. R. Statist. Soc.* 37, 149–192.
- DAVYDOV, Y., ZITIKIS, R., 2004. Convex rearrangements of random elements. *Fields Institute Communications* 44, 141–171.
- GASTWIRTH, J. L., 1971. A general definition of the Lorenz curve. *Econometrica* 39, 1037–1039.
- GOLDIE, C. M., 1977. Convergence theorems for empirical Lorenz curves and their inverses. *Adv. Appl. Prob.* 9, 765–791.
- HOEFFDING, W., 1953. On the distribution of the expected values of the order statistics. *Ann. Math. Statist.* 24, 93–100.
- KEYS, A., FIDANZA, F., KARVONEN, M. J., KIMURA, N., TAYLOR, H. L., 1972. Indices of relative weight and obesity. *Journal of Chronic Diseases* 25, 329–343.
- KINGMAN, J. F. C., 1968. The ergodic theory of subadditive stochastic processes. *J. R. Statist. Soc.* 30, 499–510.
- KOVALEVSKII, A., 2013. A regression model for prices of second-hand cars. *Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control*, 124–128.
- MACNEILL, I. B., 1978. Limit processes for sequences of partial sums of regression residuals. *Ann. Prob.* 6, 695–698.
- QUETELET, A., 1832. *Recherches sur le poids de l'homme aux different âges*. *Nouveaux Memoire de l'Academie Royale des Sciences et Belles-Lettres de Bruxelles*, VII.
- STUTE, W., 1997. Nonparametric model checks for regression. *Ann. Statist.* 25, 613–641.